

# КАРТИРОВАНИЕ НАУКИ И ТЕХНОЛОГИИ, ПРОГНОЗ РАЗВИТИЯ



М. А. Акоев  
Заместитель  
директора Центра  
мониторинга на-  
уки и образования  
Уральского феде-  
рального универси-  
тета имени первого  
Президента России  
Б. Н. Ельцина.

## MAPPING SCIENCE AND TECHNOLOGY, FORECASTING RESEARCH AND DEVELOPMENT

DOI 10.15826/B978-5-7996-1352-5.0007

*The chapter explores visualization of research documents relationships through subject, semantic, chronological and geographical mapping. Examples from different citation indexes and patent databases are used. Methods of analyzing a selection of documents using such visualizations are demonstrated. Possibilities and limitations of R&D forecasting are examined. The author highlights the ways to improve research productivity, and identify new promising directions for research.*

*В данной главе рассматриваются вопросы визуализации взаимосвязей документов в тематическом, семантическом, хронологическом и географическом представлениях на примерах публикаций из различных указателей цитирования и патентных баз. Описаны методы проведения анализа набора документов с использованием такой визуализации. В главе рассмотрены вопросы ограничений в прогнозировании научно-технического развития и аспекты повышения качества существующих научных направлений и развития новых направлений в организации.*

При переходе от наукометрического анализа отдельных ученых, коллективов и организаций к их сравнению аналитик сталкивается с необходимостью выявлять закономерности в больших объемах информации. Использование приемов визуализации данных позволяет сократить затраты на нахождение зависимостей и в дальнейшем — на представление результатов анализа. Помимо построения диаграмм к ним относится также и представление сложноорганизованной информации в формате карт

(картирование), на которых можно показать связи анализируемых элементов. С вопросами картирования наукометрической информации тесно связаны задачи построения прогноза развития науки и технологии и задачи повышения качества существующих и развития новых направлений исследований в научных организациях.

### 4.1. Картирование науки и технологии

Картирование, являясь одним из приемов наглядного представления данных, подчиняется общим

принципам визуализации, т. е. компактному представлению большого объема информации, делающему заметными найденные в ходе анализа закономерности<sup>1</sup>. Информация о типах и уровнях наукометрического анализа с примерами названий визуализаций приведена в табл. 14, построенной на основании классификации, приведенной в [Börner, Polley, 2014, p. 7]. В данном разделе для каждого типа из табл. 14 мы рассмотрим характерные примеры анализа с целью показать возможности и ограничения конкретных типов визуализации в применении к наукометрическим исследованиям. Каждый из представленных типов наукометрического анализа отвечает на один из вопросов о предмете исследования (вопросы приведены в скобках). На практике чаще всего встречается комбинация разных типов анализа, например, карта распределения публикационной активности организаций РАН будет одновременно отвечать на вопросы: где? с кем? и когда?

В настоящий момент наиболее исчерпывающие сведения о приемах научной визуализации содержатся в руководстве [Börner, Polley, 2014], а также в открытом онлайн-курсе тех же авторов [IVMOOC, 2014]. Примеры визуализаций, представленные в вышеуказанном руководстве и курсе, построены с использованием интегрированной платформы анализа и визуализации данных Sci<sup>2</sup>

<sup>1</sup>Проект «Карта российской науки» Министерства образования и науки РФ (<http://www.mapofscience.ru/>) является информационной системой для регулярно и автоматического обновления базы данных ученых и организаций, включая показатели их деятельности, осуществления статистического анализа научно-исследовательской активности и обеспечения основы для создания аналитических материалов о состоянии российского сектора научных исследований и разработок. В «Карте российской науки» использованы отдельные методы картоирования, представленные в данном разделе.

Tool, разработанной в Cyberinfrastructure for Network Science Center Университета Индианы [Sci<sup>2</sup> Tool, 2010]. Sci<sup>2</sup> Tool представляет собой модульный набор программных средств для изучения и обработки наукометрической информации, который поддерживает проведение всех типов наукометрического анализа на всех уровнях, описанных в табл. 14. Платформу Sci<sup>2</sup> Tool можно рекомендовать для изучения приемов визуализации и дальнейшей работы, при этом необходимо учитывать, что некоторые возможности по импорту и анализу данных ориентированы на данные о науке США (например, данные о грантах NSF, привязанные к формату даты США).

Примерами статистического анализа и построения профилей авторов служат визуализации в форме разнообразных диаграмм и таблиц данных. Систематическое введение в визуализацию с использованием диаграмм можно получить, обратившись к книге [Желязны, 2009] и учебнику [Яу, 2013]. Книга [Желязны, 2009] представляет приемы использования диаграмм для коммуникаций в процессе поддержки принятия решений. В учебнике [Яу, 2013] подробно описаны технические аспекты подготовки данных для визуализации. Кратко остановимся на важных аспектах визуализации, проиллюстрировав их на примере визуализации диаграммы и таблицы исходных данных.

Визуализация числа элементов в количестве большем, чем можно удерживать в оперативной памяти, нецелесообразна. Объем оперативной памяти человека определяет «магическое» число семь плюс-минус два («кошелек Миллера»). Сравните представление одной и той же информации на рис. 36а и 36б (см. с. 187). В качестве сравниваемых элементов выступают десять

Таблица 14

**Соотношение типов наукометрического анализа (в скобках указан вопрос, идентифицирующий субъект) и уровней наукометрического анализа**

| Тип наукометрического анализа                      | Уровень наукометрического анализа  |  |   |
|--|--|--|---|
|  | Микро/<br>Индивидуальный<br>(1–100 записей)  | Промежуточный/<br>Локальный<br>(101–10 000 записей)                        | Макро/<br>Глобальный<br>(более 10 000 записей)          |
| Статистический анализ / Построение профиля (Кто?)* | Экспертные данные о персональной научной продуктивности  | Исследовательские центры, университеты, предметные области или государства | Все публикации США или все публикации в мире            |
| Геопространственный анализ (Где?)                  | Индивидуальная траектория перемещения исследователя  | Распределение мест работы авторов публикаций в научном журнале             | Распределение публикационной активности организаций РАН |
| Темпоральный анализ (Когда?)                       | Динамика получения индивидуальных грантов  | Развитие тематик институтов отделения физических наук РАН за 20 лет        | 113 лет физических исследований                         |
| Тематический анализ (Что?)                         | Ключевые слова, указанные в записях о цитируемых источниках в публикациях по теме исследования | Тематики публикаций институтов отделения химии и наук о материалах РАН     | Карты ключевых слов и тематики исследований РАН         |
| Сетевой анализ (С кем?)                            | Сеть организаций, ученые которых выполняют работы в соавторстве с данной научной группой       | Сеть соавторства между исследователями по анализируемой тематике           | Ключевые компетенции РАН                                |

\*В скобках указан вопрос, идентифицирующий субъект. В ячейках таблицы приведены примеры названий визуализаций. Таблица основана на классификации, приведенной в [Börner, Polley, 2014, p. 7], с изменением примеров.

тематик, а не отдельные характеристики в ячейках таблицы. Единственное преимущество диаграммы перед табличным представлением — это возможность визуально оценить степень отличия абсолютных значений, сравниваемых элементов сразу по нескольким показателям; например, число

публикаций в первых трех областях приблизительно одинаково, но они отличаются по нормализованной цитируемости.

Визуализация как метод представления информации не может компенсировать отсутствие или низкое качество исходных данных. Также визуализация не может вы-

явить закономерности, которые не обнаруживаются количественными методами анализа. Визуализация делает закономерности в предварительно собранных и обработанных данных более доступными для восприятия. Визуализация, по сути, представляет собой краткое обобщение результатов анализа, которое позволяет понять контекст при переходе к данным, лежащим в основе анализа. Доверие к представленным в форме визуализации результатам анализа будет высоким, если обеспечивается верифицируемость результатов анализа, т. е. получение всех элементов представляемого отчета может быть прослежено до первичных данных и методик их обработки.

Пример визуализации, представленный на рис. 36а и 36б (см. с. 187), иллюстрирует необходимость обеспечения верифицируемости данных при принятии решений. Диаграмма была подготовлена для принятия решений о выборе мероприятий по развитию десяти анализируемых направлений (названия тематик для дальнейшего обсуждения несущественны, публикации были отобраны в ручном режиме с привлечением экспертов в предметной области и представляют узкие тематические направления). Для всех анализируемых областей характерен уровень цитируемости ниже, чем среднемировой, что позволяет для всех научных групп сформулировать задачу повышения видимости их научного результата в мире. Для научных групп, которые представляют тематику Subject 10, только из анализа представленных данных можно было бы сделать вывод, что при выделении дополнительного финансирования можно ожидать роста качественных показателей публикаций при сохранении количества публикаций. К данному выводу можно прийти исходя из того, что ученые,

опубликовавшие анализируемые работы, при высоких темпах роста смогли сохранить самый высокий уровень востребованности своих результатов в анализируемых предметных областях (косвенно об этом свидетельствует значение нормализованной цитируемости публикаций). Однако при обращении к первичным данным о цитируемости публикаций данной предметной области выявляется высокий процент самоцитирования. Из этого следует, что на фоне увеличения числа публикаций по тематике группа в меньшей степени ориентирована на привлечение внешнего внимания к своим работам. Принятие решений по выделению ресурсов для данной группы требует дополнительного анализа с привлечением экспертов в этой предметной области.

Возможность визуализации данных зависит от наличия готовых инструментов построения визуальных представлений. Только опыт регулярных созданий и представлений визуализаций при наличии инструментов и подготовленных для визуализации данных гарантирует создание оптимального, с точки зрения восприятия, типа визуального представления результатов. Многие типы визуального представления результатов, обобщающие опыт многих пользователей, встроены в библиометрические базы данных или аналитические надстройки над ними (подробнее про библиометрические инструменты см. в разделе 2.1). Однако при проведении исследований с одновременным использованием разных типов анализа (табл. 14) штатных средств анализа обычно бывает недостаточно. Приходится прибегать к использованию готовых инструментов или создавать свои инструменты визуализации под конкретный вид представления или подготовку данных для визуализации. Обзор готовых

средств визуализации приведен в [Science mapping..., 2011].

Картирование — одна из сложнейших форм визуализации как по трудоемкости процесса подготовки данных, так и по восприятию и интерпретации результата потребителем. Трудоемкость подготовки данных определяется необходимостью, с одной стороны, указать координаты на карте для каждого отображаемого элемента (публикации, исследователя и т. п.), а с другой — получить «картографическую» основу, полезную для целей анализа и отражающую представление о реальности. Например, в случае геопространственных визуализаций «картографическая» основа отражает реальный мир и распределение элементов в географическом пространстве, а в случае тематических карт или схем — соотнесение тематик, которые отражают науку ровно в той степени, в которой ее отражают исходные данные, использованные для построения основы карты. Существенное отличие тематических «картографических» основ от географических в том, что мы не можем на момент построения знать все тематики исследований с учетом будущих результатов исследований. Тематические карты не обладают предсказательной силой «белых пятен» географических карт.

Основная ценность картографического способа визуализации данных заключается в возможности продемонстрировать распределение элементов в условном пространстве карты, показывая расстояния между ними. Зная распределение элементов в пространстве карты, мы можем находить в окрестности выбранного элемента все существующие элементы, которые считаются близкими, например тематически связанные публикации одного временного периода. Именно возможность определить окрестности выбранного элемента делают

картографические способы отображения полезными для анализа, так как позволяют найти эталонный набор элементов для сравнения. Отметим, что поиск окрестностей элемента не является следствием их картографического отображения. Обе возможности — как поиск, так и отображение — являются следствием моделей, лежащих в «картографических» основах. Обзор подходов к выбору моделей для построения «картографических» основ представлен в [Mapping research..., 2008].

#### **4.1.1. Геопространственный анализ (где?)**

В качестве примера геопространственного анализа построим карту мест работы авторов, указываемых в аффилиации в журнале ACM Transactions on Database Systems (TODS). Публикации в научном журнале относятся к узкой тематике, исключения составляют мультидисциплинарные журналы (подробнее обсуждение классификаций журналов см. в разделе 3.7). Следовательно, по аффилиации авторов можно отследить, где расположены научные центры, в которых выполняются работы по узкой тематике. Результат анализа представлен на рис. 37 (см. с. 188). Информация о координатах каждого места работы была получена геокодированием адресной информации из аффилиации авторов в библиографических записях. Геокодирование — это автоматизированный процесс сопоставления почтовых адресов с координатами на карте (в данном анализе использовалась служба Geocoding API Google<sup>2</sup>). За анализируемый период, с 2008 г. по первый выпуск 2014 г., в журнале была опубликована 181 работа, что составляет порядка 21 % всех публикаций в журнале с момента его создания (журнал

<sup>2</sup><https://developers.google.com/maps/documentation/geocoding/> (дата обращения: 25.06.2014).

выходит четыре раза в год). Данные работы написали 437 авторов, которые указали 420 текстуально разных вариантов аффилиации (некоторые авторы указали более одной аффилиации). Из представленной на рис. 37 (см. с. 188) карты видно, что авторы за небольшим исключением сконцентрированы в Европе и США.

Процесс, использующий автоматическое геокодирование, не всегда возвращает правильные координаты. Например, отметка в Тихом океане у берегов Чили является результатом геокодирования следующей аффилиации: «Departamento de Ciencias de la Computación, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Chile» (при повторной обработке в июле 2014 г. геокодирование данной строки возвращало координаты в г. Сантьяго). Также при геокодировании необходимо учитывать, что не все аффилиации могут быть корректно найдены, например, строка «Microsoft Research Asia, Beijing, China» может быть распознана по координатам географического центра Пекина, а Microsoft вернет неопределенное значение в силу множества подходящих вариантов. При геокодировании адресов необходимо учитывать, что, например, для действительных членов РАН геокодирование аффилиации может возвращать координаты адреса: г. Москва, Ленинский просп., д. 14.

Применение созданных при разработке геоинформационных систем (ГИС) и активно применяемых в географии методов геопространственного анализа к наукометрическим данным находится в начальной стадии освоения. Систематическое применение методов геопространственного анализа требует использования специализированных пакетов ГИС. Для начального знакомства с темой можно порекомендовать

обратиться к форуму GIS-Lab<sup>3</sup> и в дальнейшем перейти к чтению специализированной литературы.

#### **4.1.2. Темпоральный анализ (когда?)**

Исторически темпоральный анализ как инструмент анализа наукометрических данных появился первым, что связано с использованием SCI для установления научных приоритетов в работах 1940-х гг. Первую публично-доступную программу визуализации публикаций на временной шкале с указанием связей между ними (HistCite) создал Юджин Гарфилд [HistCite, 2012]. Пример визуализации с использованием HistCite можно увидеть в статье Ю. Гарфилда «Эволюция Science Citation Index», включенной в данное руководство.

При использовании инструмента темпорального анализа в визуализации возможны два варианта представления времени: либо в качестве одной из осей, либо в последовательности анимированных изображений. Второй способ применяется для визуализации структур, которые требуют двумерного отображения, например, динамика во времени и географическое распределение публикационной активности авторов журнала. Визуализировать время для сложных структур также можно, используя цветовое кодирование времени, если это уместно. Однако этот способ менее распространен, так как затрудняет чтение карт, перегружая их визуальной информацией.

Самым простым примером темпорального анализа является столбчатая диаграмма распределения числа публикаций во времени. Для демонстрации возможностей темпорального анализа, требующего отображения на карте, приведем распределение во времени связей между кластерами публикаций авторов, прямо или косвенно

<sup>3</sup><http://gis-lab.info>

цитирующих самую раннюю из самых цитируемых в наукометрии публикаций математика А. Дж. Лотки [Lotka, 1926] (рис. 38, см. с. 189). Визуализация выполнена в инструменте CitNetExplorer, развивающем идеи инструмента HistCite [CitNetExplorer, 2014]. CitNetExplorer позволяет визуализировать большие массивы публикаций и кластеризовать их, скрывая несущественные детали [Eck, Waltman, 2014].

Визуализация, представленная на рис. 38 (см. с. 189), была выполнена на основе публикаций за период 1945–2014 гг., полученных из Web of Science CC для 14 журналов, перечисленных в табл. 15. Всего было найдено 37989 публикаций и 232202 ссылки на них в анализируемом массиве. Визуализация воспроизводит пример, приведенный в лекции Н. Ван Эйка (N. Van Eck) о возможностях CitNetExplorer на конференции STI в 2014 г.

Визуализация кластеров анализируемых публикаций позволяет увидеть все ключевые фамилии авторов, оказавших наибольшее влияние на развитие наукометрии и упоминавшихся в предыдущих главах. Также на представленном рис. 38 (см. с. 189) можно увидеть связи группы публикаций по наукометрии (они выделены зеленым

цветом) с группой публикаций по информационному поиску (выделены синим цветом), например, с публикациями Г. Сэлтона (G. Salton) и с группой публикаций автора концепции тройной спирали Г. Ицковица (H. Etzkowitz) [Ицковиц, 2010] (выделены фиолетовым цветом).

#### 4.1.3. Сетевой анализ (с кем?)

Для исследования науки как процесса научной коммуникации можно использовать инструменты сетевого анализа, развитые в дисциплине Social Network Analysis (SNA). Сетевой анализ позволяет выявить закономерности в совместных работах отдельных авторов и научных организаций. Однако возможности применения сетевого анализа для наукометрических исследований реализуются в полной мере только при наличии машиночитаемой информации о научных связях ученых, о полученных ими результатах и применении специализированных инструментов. В качестве инструмента для знакомства с возможностями сетевого анализа рекомендуется использовать Sci<sup>2</sup> Tool [Sci<sup>2</sup> Tool, 2010]. В качестве инструмента анализа и визуализации сетей большого размера можно порекомендовать бесплат-

Таблица 15

#### Журналы по наукометрии с 1945 г., индексируемые в Web of Science CC

|   |   |
|---|---|
| American Documentation                                | Journal of Information Science                                    |
| The Annual Review of Information Science & Technology | Journal of Informetrics   |
| Aslib Proceedings                                     | Journal of the American Society for Information Science           |
| Information Processing & Management                   | Journal of the Association for Information Science and Technology |
| Information Scientist                                 | Research Evaluation   |
| Information Storage and Retrieval                     | Research Policy   |
| Journal of Documentation                              | Scientometrics  |

ную программу Pajek [Pajek, 1996]. Однако для проведения сложных видов анализа необходимо владеть компетенциями для работы со специализированными продуктами и освоить навыки программирования. Если у аналитика отсутствуют необходимые компетенции программирования, то для преобразования и подготовки данных для анализа в проект нужно привлекать программиста. Однако аналитику желательно самостоятельно освоить основы программирования в одном из онлайн-курсов анализа и обработки данных<sup>4</sup>. Для большинства задач подготовки данных требуются минимальные навыки программирования. Дополнительным плюсом от освоения программирования будет улучшение взаимопонимания с программистами, которым вы ставите задачи.

В качестве примера сетевого анализа построим сеть соавторства (рис. 39, см. с. 190) для публикаций четырех авторов: Стенли Васерман (S. Wasserman), Юджин Гарфилд (E. Garfield), Алесандро Веспиньяни (A. Vespignani) и Альберт-Ласло Барабаши (A.-L. Barabasi), 361 запись, период — 1955–2007 гг., источник — Web of Science CC. Построение выполнено с использованием Sci<sup>2</sup> Tool, исходные данные визуализации входят в комплект поставки данного инструмента. Методика построения сетей описана в [Börner, Polley, 2014, p. 204–206]. Итак, соавторы образуют четыре четко выраженных кластера. Два кластера публикаций, Стенли Васермана и Юджина Гарфилда, не связаны между собой и с остальными кластерами и, таким образом, образуют изолированные «острова» соавторства. Выявление изолированных «островов» в сетях соавторства для публикаций по одной тематике может

свидетельствовать как о возможности наладить новые научные связи, так и о неверном отнесении публикаций к тематике. Выполняя анализ только на основе данных о соавторстве, без привлечения дополнительной информации, например о ко-цитировании, можно прийти к неверным выводам, так как перечисление в списке авторов не несет информации о вкладе каждого из них в написанную статью.

Визуализации сетевого анализа на промежуточном и макроуровнях (табл. 14) становятся нечитаемыми, так как содержат множество узлов с одной-двумя связями, загромождающими построенную карту. Концентрация связей позволяет визуально подчеркнуть центры анализируемой активности, однако затрудняет общее восприятие и делает метки нечитаемыми. Для представления карт, построенных по результатам сетевого анализа, рекомендуется сокращать концентрацию связей, используя вместо них цветовую раскраску, например тепловую карту (heatmap), или используя в качестве дополнительного параметра кольцо вокруг узла с данными.

Сетевой анализ не замыкается только на анализе библиометрических данных. Ученые, активно применяя современные коммуникационные возможности при работе с информацией, дают аналитикам возможность использовать при анализе данные, сбор которых ранее был затруднен или невозможен. В качестве примера можно привести сетевой анализ и визуализацию истории переходов читателей между статьями на сайтах издательств [Clickstream Data..., 2009]. За два года, с 2007-го по 2008-й, была собрана информация о более чем одном миллиарде обращений к страницам полнотекстовых ресурсов, данные о переходах и идентификации пользователей

---

<sup>4</sup><https://www.coursera.org/courses?categories=stats>



представили издательства. Анализ, выполненный в статье, базируется на предположении, что большинство переходов на страницы издательств с целью получения полного текста были предприняты для чтения запрашиваемой работы как минимум с целью ознакомления. Полученная в результате анализа карта (рис. 40, см. с. 191) представляет данные о связях между журналами на основе анализа динамики перехода читателей между страницами изданий. Связи между журналами отражают статистически значимое число переходов в процессе чтения.

На карте сетевого анализа можно увидеть ожидаемую концентрацию переходов между журналами одной тематики и переходы между журналами традиционно связанных тематик. Однако из карты следует «неожиданный» результат — в центре карты область социально-гуманитарных журналов оказалась связана с областями естественных, технических и медицинских наук, что не соответствует картам, которые строятся только на основании данных о цитировании публикаций. В качестве дополнительного примера сетевого анализа в применении к наукометрии гуманитарных дисциплин можно привести исследование Р. Коллинза [Collins, 2000]. В рамках данного исследования были выявлены и проанализированы связи между философами на протяжении более чем двух тысяч лет и было показано, что в философии наблюдаются процессы, идентичные тем, которые мы видим в естественных науках. В качестве примера выбора источников данных для проведения сетевого анализа внутри организации можно обратиться к диаграмме, представленной в разделе 1.5.

#### **4.1.4. Тематический анализ (что?)**

Ответы на вопросы «Что является предметом исследований?» и «Как связаны

предметы исследований между собой и с практикой?» — это священный Грааль наукометрии. Основным препятствием в ответах на оба вопроса является научный процесс, который ориентирован на порождение компактного представления новых знаний для нужд практики, а не для облегчения наукометрических исследований (подробнее см. в разделе 1.2). Предмет исследований надежно идентифицируется только комбинацией трех элементов: текстов с представлением результатов исследований, контекста привлеченных для аргументации ссылок на факты и имплицитных знаний в голове сообщества исследователей и практиков, работающих над предметом исследований. Современные методы, используемые для идентификации предмета исследований, являются грубым приближением, ориентированным на сокращение числа документов, которые необходимо просмотреть и прочитать в процессе поиска информации. Достигнутое качество аппроксимации определяется прежде всего потребностями поиска информации, а не проведением наукометрических исследований. Для записей баз данных о публикациях считается достаточным, если классификация статьи базируется на журнальной классификации. Единственное неудобство возникает при классификации статей из мультидисциплинарных журналов (подробнее обсуждение классификаторов см. в разделе 3.7).

Возможности поиска информации в базах данных публикаций, основанных на словах в названии статей и источнике, в котором опубликован текст, могут быть дополнены поиском по ключевым словам. Ключевые слова способствуют уточнению тематики публикации, но не подменяют содержимое публикации. Два ограничения препятствуют широкому использованию ключевых

слов для идентификации предмета исследования. Это, во-первых, необходимость указывать достаточное число ключевых слов для полноты идентификации тематики статьи, а во-вторых, проблема унификации и синонимии с целью повышения точности поиска. Для многих российских статей по экономике одно из наиболее популярных ключевых слов — это «экономика». Преодолеить подобные ограничения может механизм приписывания ключевых слов на основе информации из библиографии, указанной в публикации, например инструмент KeyWords Plus, используемый в Web of Science CC. К сожалению, подобные механизмы не могут полностью заменить работу автора текста или издателя при генерации поисковых образов документов для всех записей в базе.

Несмотря на ограничения, присущие ключевым словам, их можно использовать прежде всего для проведения поиска, а также ориентироваться на них при выявлении тематик исследований. Для примера визуализируем по ключевым словам изменение тематики публикаций авторов из Казахстана с 2004 по 2014 г. (см. рис. 41 — с. 192), используя в качестве инструмента облако ключевых слов Wordle<sup>5</sup> [Participatory..., 2009]. Число публикаций авторов из Казахстана, отраженных в Web of Science CC, за указанный период удвоилось (периоды сравнения — пять лет). Облако ключевых слов строится по значению поля KeyWords Plus, которое указано менее чем в половине записей. Размер ключевого слова отражает долю публикаций в общем числе публикаций. Из динамики ключевых слов и факта просмотра соответствующих им публикаций можно сделать вывод, что тема моделирования и астрофизики в Ка-

захстане на подъеме, а интерес к экологии водных ресурсов снизился.

На представленной визуализации можно увидеть и основное ограничение инструмента облака ключевых слов, а именно отсутствие унификации ключевых слов. Слово «модель» в единственном и множественном числе одновременно присутствует на диаграмме, и без обращения к исходным публикациям выяснить контекст применения этих понятий невозможно. Использовать визуализацию в виде облака ключевых слов желательно только на массиве, относящемся к одной предметной области.

Альтернативой облаку ключевых слов, лишенной их недостатков, является инструмент картирования на основе модели векторного пространства (Vector space model, VSM). Обзор современных методов приведен в [An approach to..., 2009]. Примером использования модели векторного пространства является построение патентных ландшафтов. Его применение возможно, если все записи на публикации в базе снабжены ключевыми словами, обеспечивающими полноту и точность описания публикации для поиска. В качестве примера такой базы данных можно привести международную реферативную патентную базу данных Derwent World Patents Index, размещенную на информационной платформе Thomson Innovation Thomson Reuters. Каждая запись о патенте снабжена описанием патентной формулы в форме ключевых слов, составленным экспертами вручную (в дополнение к патентной формуле, описанной в тексте патента), которое позволяет идентифицировать существенное сходство патентов, выданных в разной юрисдикции и разным правообладателям. Единицей анализа в Derwent World Patents Index выступает патентное семейство — пул

<sup>5</sup><http://www.wordle.net/>

патентов, обладающих идентичной патентной формулой.

Пример патентного ландшафта приведен на рис. 42 (см. с. 193). В основе построения патентного ландшафта лежит метод измерения расстояний между двумя патентными семействами. Расстояние между ними определяется на основании числа ключевых слов описания патентной формулы, на которое отличаются два описания. Патентный ландшафт строится по пулу предварительно отобранных патентов, т. е. не существует единой картографической основы, на которую можно нанести патенты и увидеть их глобальное взаиморасположение. Если у двух патентных семейств, для которых вычисляется расстояние, нет общих ключевых слов, то расстояние вычисляется как минимальный путь из всех возможных путей, образуемых пулом анализируемых патентов, включающих оба семейства, между которыми вычисляется расстояние.

С помощью цветов на патентном ландшафте кодируются названия тематических областей. Области не изолированы между собой, например, область гражданского строительства в анализируемом пуле граничит с материаловедением, информационно-телекоммуникационными технологиями, электроникой и полупроводниками. Изменением оттенка и изолиниями на ландшафте показаны области с высокой патентной активностью, условные вершины на карте подписаны для ориентации. Так же как и остальные виды картографической визуализации, патентные ландшафты не позволяют принимать решения без дополнительного анализа патентных семейств. Например, условная низина на ландшафте, в которой у организации наблюдается концентрация патентов, может служить указанием на потенциально пер-

спективную область получения патентов для данной организации, но также может свидетельствовать о том, что область осталась в развитии. Возможность построения патентных ландшафтов подтверждает возможность выявления сходных патентов и их последующего анализа.

Основная ценность библиографических баз данных, обеспечивающих точность и полноту поиска по предмету исследований, заключается прежде всего в возможности снизить затраты специалистов на поиск публикаций. Если информация к каждой записи добавляется единообразно и в машиночитаемом виде, то возникает возможность найти близкие по тематике записи и использовать их для расчета средних показателей, которые могут быть референтными для сравнения показателей научной результативности как для отдельных ученых, так и для научных коллективов (подробнее см. в разделах 3.6 и 3.9). Подобные базы позволяют автоматизированно выявлять закономерности в данных о публикациях и на их основании косвенно судить о динамике приоритетов развития науки в ретроспективном режиме.

Эталонным примером подобных баз выступали реферативные журналы ВИНТИ, в которых для каждой отраженной в них записи приведены сведения о классификации записей, ключевых словах и написан реферат специалистом в предметной области, который дает представление о содержимом работы. Создание и поддержание в актуальном состоянии подобных баз данных требует привлечения большого числа квалифицированных специалистов. Недостатком подхода к созданию баз «с ручным» реферированием является также и задержка с обновлением информации в базе, так как от момента появления публикации до

появления реферата на нее проходит полгода и более. Эффективно работать над развитием научного направления с подобными задержками затруднительно. Однако для задач мониторинга трендов развития науки и технологии подобный подход эффективен.

В некоторых областях науки возможно альтернативное решение по идентификации предметов исследований, снижающее зависимость от человеческого фактора при создании поискового описания документов. Например, в органической химии активно используются подходы по идентификации предмета исследований на основании структурной формулы соединений, упоминаемых в статье. В качестве подобной базы данных можно привести базу Chemical Abstracts (SciFinder), выпускаемую Американским химическим обществом. Используя единый подход к описанию структурных формул органических соединений, можно добиться однозначности в идентификации предмета исследований, но без обращения к тексту статьи невозможно получить однозначное понимание контекста, в котором упоминается данное вещество. Как и в примере с реферативными журналами ВИНТИ, в Chemical Abstracts создание поискового описания требует квалифицированного труда, однако поскольку для идентификации соединений в тексте не требуется полностью читать статью, обновление базы происходит без существенных задержек. Можно привести в качестве примера несколько баз, использующих аналогичные подходы к созданию описаний, основанных на возможности идентификации заболеваний по международной классификации болезней (International Statistical Classification of Diseases and Related Health Problems) или индекса художественных произведений.

Использованию идентификации как инструмента для поиска всех научных публикаций зачастую препятствует отсутствие надежных методов однозначной идентификации предметов исследований и высокие затраты на создание и актуализацию подобных баз. Затраты на создание баз данных для областей с разработанными методами идентификации в будущем могут быть снижены в результате развития методов автоматического извлечения информации, однако универсальных подходов к созданию методов однозначной идентификации на данный момент не создано.

Компромисса между скоростью создания поискового аппарата, его качеством, обеспечивающим полноту и точность поиска, и стоимостью актуализации библиографических баз данных сумели достичь при использовании такого инструмента, как цитирование — естественное для научной публикации средство указания на факты, привлекаемые для доказательства утверждений (обсуждение связи фактов и ссылок приведено в [Латур, 2013, с. 64–106], также см. раздел 3.3). Science Citation Index как инструмент в первую очередь сделал доступным перспективный поиск информации, а также позволил использовать информацию, аппроксимирующую предмет исследований публикации для автоматического анализа информации.

Использование информации о том, с какими свойствами цитировались публикации в анализируемом массиве публикаций, позволяет наглядно показать, как связаны группы публикаций. В качестве примера на рис. 43 (см. с. 194) приведена карта публикаций авторов СПбГУ, группировка работ выполнена по журналам, в которых они были опубликованы. Для визуализации использована программа VOSviewer [VOSviewer,

2013], позволяющая представлять результаты анализа методами «библиографического сочетания», ко-цитирования и соавторства [Eck, Waltman, 2010]. Основы перечисленных методов анализа рассмотрены в разделе 3.4. Визуализация методом «библиографического сочетания» позволяет судить о том, как связаны публикации по совместно используемым источникам информации: например, для карты на рис. 43 (см. с. 194) связь осуществляется по ссылкам на публикации в одних и тех же журналах. Классификация тематики публикаций закодирована цветом и соответствует классификации, используемой в Web of Science CC (см. раздел 3.7). На карте видно, что публикации по математике, физике, геологии и химии образуют последовательность сильно связанных звеньев одной цепи. К этим статьям примыкают публикации по астрономии и отдельные публикации по биологии. При этом основная часть биологических публикаций образует отдельный «архипелаг», что может говорить о том, что биологическая тематика и остальные исследования не имеют существенных общих предметов изучения в университете. Без проведения дополнительного анализа публикаций невозможно сделать вывод о причинах подобной обособленности.

Построение визуализации связей между публикациями на макроуровне способствует пониманию связей между научными дисциплинами. Часто подобным визуализациям дают название – карта науки. Такое громкое название, скорее, отражает амбиции авторов. В качестве примера на рис. 44 (см. с. 195) показана карта науки, построенная методом ко-цитирования. Принцип, определяющий размер и цветовое кодирование кругов, отражающих число публикаций в каждом журнале в выбранный год, тот же, что и на пре-

дыдущем рисунке. Взаимное расположение кругов на диаграмме определяется на основании числа цитирований идентичных наборов публикаций в работах, опубликованных в каждой паре журналов. В основе анализа, базирующегося на ко-цитировании, лежит предположение, что статьи связаны не только общим предметом исследований, но и контекстом, в котором происходит исследование общего предмета, если есть более чем одна общая цитируемая работа. Необходимо обратить внимание на отличие связей, существующих между научными дисциплинами в представленной карте науки, от визуализации связей между журналами, построенных на основании анализа истории переходов между журналами на сайтах издательств (см. рис. 40 – с. 191). Данный факт можно интерпретировать следующим образом: ссылки на социальные и гуманитарные исследования не указываются в работах по дисциплинам, связанным на карте чтения как неприемлемые для «строгих» дисциплин. Однако возможны и другие объяснения, например, что представители остальных дисциплин читают социальные и гуманитарные исследования для общего развития. Однозначно ответить на вопрос о причинах противоречий между картами, отражающими два взгляда на науку, без дополнительных исследований не представляется возможным.

После визуализации связей и взаимного расположения журналов на основании анализа ко-цитирования их публикаций можно перейти к построению классификаций журналов, притязающих на естественность по способу своего построения. Классификация журналов, построенная на основе ко-цитирования публикаций, отражает связи между классами журналов, определяемых на основании общности предмета исследований. Примером подобной классифи-

кации является карта науки, созданная для University of California, San Diego (UCSD) и представленная на рис. 45 (см. с. 196). Методика построения базируется на анализе ко-цитирования публикаций за три периода с 2001 по 2010 г. по данным Web of Science CC и уточнена по данным Scopus. Детали методики построения и история построения карты описаны в [Design and Update..., 2012]. Итоговая классификация содержит 13 категорий и 554 подкатегории. Увеличение дробности классификации и использование классов ко-цитирования в качестве основания для выделения приводит к появлению интересных классов (табл. 16). Названия категорий свидетельствуют о том, что публикации в журналах, которые входят в представленные подкатегории, существенно изолированы от остальных публикаций в категориях. Отметим, что в последующих обновлениях классификации к подкатегориям «российская математика» и «российская химия» добавилась и «российская физика».

Картирование с использованием готовой тематической топоосновы, например пред-

Таблица 16

### Примеры категорий классификации UCSD

| Номер UCSD | Подкатегория                  | Категория    |
|------------|-------------------------------|--------------|
| 393        | Mathematical Science (Russia) | Math&Physics |
| 491        | Chemistry (Russia)            | Chemistry    |

Источник [Design and Update..., 2012]; данные опубликованы на сайте <http://sci.cns.iu.edu/ucsdmap/>

ставленной на рис. 45 (см. с. 196), позволяет сравнить степень специализации анализируемого массива публикаций со всеми публикациями мира, а фиксированное расположение элементов классификации на диаграмме позволяет быстро читать и срав-

нивать карты, построенные для разных «пулов» публикаций. На рис. 46 (см. с. 197) представлен результат анализа публикаций авторов из СПбГУ, которые уже анализировались на рис. 43 (см. с. 194). Сравнивая два метода визуализации, можно отметить, что на карте, построенной на основании информации только из анализируемых публикаций, незаметны исследования в области мозга, медицины, а также не охвачены социально-гуманитарные области. Таким образом, дополнительным достоинством метода картирования на основании готовой тематической топоосновы является возможность увидеть относительно небольшие группы исследований, которые в случае применения другого метода визуализации теряются на фоне публикаций, относящихся к тематикам с самым большим числом публикаций.

Описанный выше метод, лежащий в основе построения естественной классификации журналов на основании ко-цитирований, можно применить и для классификации отдельных публикаций. Классификация отдельных публикаций позволяет надежно определить категорию, к которой относится публикация, не на основании классификации журнала, а на основании тематической окрестности каждой публикации. Это позволяет преодолеть сложность в классификации статей в мультидисциплинарных журналах. Используя данный подход, можно не ограничиваться уже предложенными категориями, а строить существенно более подробную систему классификации. Однако данный метод как в отношении журналов, так и в отношении статей обладает существенным недостатком. Если мы построили классификацию на массиве данных, например, за 2001–2010 гг., то надежность классификации публикаций, написанных

существенно позднее данного периода, будет сомнительной, ведь за это время может существенно поменяться схема связей между категориями. Однако обновление классификации на основании более свежих данных приводит к тому, что элементы классификации изменяются. Изменения категорий верхнего уровня, может быть, и незначительные, но для узких категорий, используемых для классификации отдельных публикаций, становится невозможным проведение сравнений между классификациями, построенными на массивах публикаций разных лет.

Представленные в разделе техники визуализации информации о науке и технологии путем картирования наукометрической информации позволяют облегчить понимание результатов анализа, однако они не могут подтвердить существование закономерностей, которые не следуют из исходных данных. Доступность современных инструментов визуализаций позволяет с минимальными усилиями создавать большое число диаграмм, которые привлекут внимание к представлению результатов анализа, однако могут содержать дублирование уже представленной информации. В большом числе визуализаций можно скрыть недостатки проведенного анализа и сформировать искаженное понимание представленных результатов.

## **4.2. Прогнозирование и развитие**

Лица, принимающие решения при обсуждении методов картирования науки и технологии, неявно считают, что подобные методы обладают свойствами географических карт, важнейшим из которых была идентификация «белых пятен», позволяющая «планировать» научные и технологические достижения. Тематические «картографиче-

ские» основы принципиально отличаются от географических тем, что мы не можем на момент построения карты знать все потенциально возможные тематики исследований, из которых можно было бы сложить карту для прогнозирования развития науки и техники.

Одно исключение, при котором возможно плановое управление направлением развития с использованием для прогноза методов «наукометрии», представлено в форме схемы системной динамики (рис. 32а). Речь идет о догоняющем развитии науки отдельной страны. Если в мире есть технологический лидер, который обладает достаточным объемом ресурсов для проведения исследований по широкому спектру направлений, то догоняющая страна может сэкономить свои усилия на развитие собственных перспективных научных исследований за счет мониторинга изменения тематик исследований у лидера и соответственно перераспределять имеющиеся научные ресурсы для воспроизведения уже полученных результатов, «подражая» поведению научного лидера. Если у научной тематики возникает возможность прикладного применения, то подобный подход может гарантировать, что в стране найдется специалист, который разбирается в проблеме или может быстро переключиться на нее за счет того, что он занимается тематически близкой проблематикой. У предложенной схемы есть два недостатка. Во-первых, всегда будет наблюдаться отставание от лидера, так как доступные для мониторинга изменения тематик на основе видимых результатов зачастую появляются в момент, когда авторы публикации работают над развитием тематики или уже отказались от нее. Задержка возникает вследствие необходимости донесения проблематики до конкретного ис-



следователя, который может находиться вне контекста исследований. Во-вторых, в целях экономии ресурсов эксперты могут пожертвовать тематиками, развиваемыми у лидера, которые не соответствуют высокоуровневым приоритетам практической деятельности в стране. Эти тематики могут быть востребованы в дальнейшем, но в момент начала работ их необходимость неочевидна.

На схеме (рис. 32а) представлено формирование прогноза развития науки и техники на основании мониторинга результатов уже выполненных исследований. Эксперты в данной схеме играют роль оракулов при мониторинге научных результатов, получаемых в стране-лидере. В предложенной модели возможно внесение небольших корректировок в прогноз развития за счет учета потребностей развития локальных исследований (указано пунктирной стрелкой), однако эта коррекция незначительна. Неявно в схеме предполагается, что эксперты при составлении прогноза развития учитывают потребности развития приоритетных направлений техники, если это необходимо.

Обратим внимание на потенциальную проблему в рассматриваемой схеме, а именно на отрыв экспертов от практической деятельности и полное их сосредоточение на анализе входного документального потока информации для составления плана развития. Также у исследователей атрофируется навык поиска новых направлений исследований и развиваются иждивенческие настроения в отношении получения финансирования. Использование принципа, заложенного в схеме на рис. 32а, предусматривает установление такого уровня развития науки и техники, при котором развитие переходит на методы, используемые страной-лидером. Обсуждение уровней

развития науки в стране описано в [Tracking scientific..., 2014].

Схема формирования направлений развития науки и техники, реализуемая в странах-лидерах, представлена на рис. 32б. Она визуализирует процедуру, описанную в разделе 1.1. В предложенной схеме эксперты находятся в тесном контакте с исследователями, а часто и сами являются исследователями и находятся в контексте тематик исследований. Данные наукометрии привлекаются как вспомогательный инструмент для калибровки и настройки понимания экспертом цели экспертизы при формировании экспертных заключений. В результате

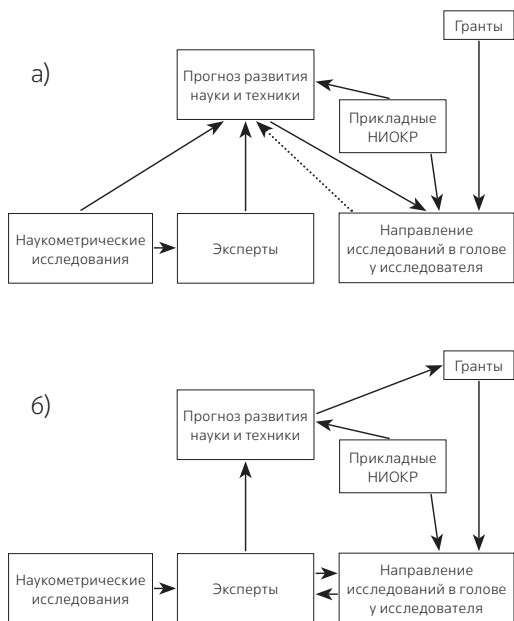


Рис. 32. Варианты управления тематикой научных исследований: а) плановое управление с использованием наукометрии для прогноза; б) схема подстройки направлений исследований, формирующая цикл управления через изменение приоритетов грантового распределения ресурсов



процесса экспертизы формируются направления развития науки и техники, которые используются при изменении приоритетов в выделении объемов ресурсов для научных исследований по тематике (на схеме представлены в форме грантов).

Если возможности применения схемы рис. 32а в развитии науки ограничены, нельзя ли применить ее для развития технологии? Если правообладатель не заинтересован в получении локального патента, то это хороший шанс получить патент в своей юрисдикции и наладить выпуск продукции по имеющимся описаниям. Однако не во всех технологических областях все продукты защищаются патентами. В исследовании инновационных продуктов, получивших в период 1977–2004 гг. ежегодную премию R&D 100 Awards журнала Research and Development, было обнаружено, что на 90,9 % продуктов не были получены патенты [Reassessing patent..., 2013]. Вопросы полноты отражения знания технологий в патентах требуют дополнительных исследований, так же как и вопрос о выявлении связи между научным и технологическим знанием (см. обсуждение в разделе 1.2).

Если схема рис. 32а не работает в рамках страны, может ли она сработать в рамках отдельной научной организации? Да, может, но только на начальном этапе развития. Например, созданный в 2009 г. Университет науки и технологии имени короля Абдулы (King Abdullah University of Science and Technology, KAUST), определив тематический план своего развития и привлекая высококвалифицированных ученых со всего мира, смог увеличить объем и качество своих научных результатов до уровня, позволившего ему войти в топ-500 Шанхайского рейтинга. Сможет ли университет продолжить свое развитие, если прекратит-

ся подпитка высокорейтинговыми учеными, покажет время.

При принятии решения о развитии новых научных направлений необходимо учитывать доступность ресурсов человеческого капитала и специфику региона, понимание которой не может быть достигнуто средствами наукометрии. Такие данные можно получить исключительно через экспертов и уже работающих в организации ученых, знакомых с местной спецификой и местной научной составляющей.

При определении направлений развития нужно выбирать тематики, которые позволят сформировать жизнеспособные научные группы. Критерии для выбора тематики, которая позволяет сформировать жизнеспособную научную группу, основываются не столько на выявленных методами наукометрии перспективных направлениях, сколько на наличии возможности создать группу ученых, которая сможет привлечь к обсуждению и использованию своих результатов коллег и практиков вне организации (см. обсуждение в разделе 1.2).

Работа научной группы по новому направлению начинается с чтения публикаций по теме исследований. Чем лучше группа исследователей знакома с уже вышедшими публикациями, тем лучше понимание членами группы контекста исследований, проводимых в мире, и тем выше шансы предложить для обсуждения новые результаты, которые отвечают на актуальные вопросы. В качестве средства ранней диагностики развития научного направления можно предложить отслеживание динамики чтения. Связь между чтением и последующими публикациями статистически значима и подтверждалась в большом числе исследований университетов и научных организаций, ориентированных на создание ново-

го знания. Динамика увеличения числа обращений к полнотекстовым ресурсам может задолго до появления первых публикаций сигнализировать о том, что процесс развития нового направления запущен.

Для членов новой научной группы также требуются и навыки представления научных результатов (см. обсуждение в разделе 1.5). Привлечение внимания к результатам своих исследований и, как следствие, повышение уровня качества научных результатов подразумевает систематическое чтение и обсуждение актуальных научных статей, поездки на семинары/конференции для общения с коллегами и получение обратной связи о потребностях от местной промышленности и общества.

Навыки систематического чтения научной литературы, обсуждения и представления новых научных результатов не исчерпывают всех навыков, необходимых исследователям в научных группах. Список универсальных навыков исследователей, на которые нужно обращать внимание при обучении новых научных сотрудников или привлечении существующих, был определен в рамках исследования мобильности научных сотрудников на основе результатов анализа факторов, способствующих присвоению степени PhD в Европейском союзе, см. табл. 17 [Transferable Skills..., 2012, p. 20].

Предложенный выше процесс формирования жизнеспособных научных групп для развития новых научных направлений в организации требует учета потребности промышленности (практики). Хорошим индикатором востребованности уже выполненных исследований в практике является регулярное выполнение научной группой прикладных работ в интересах внешнего заказчика. Однако для новых научных групп этот индикатор не является релевантным, так как

группой еще не получены результаты, которые могут привлечь практиков. Исключение составляют группы, которые создаются изначально для решения практических задач, а в дальнейшем переходят к развитию научной тематики. Для выявления и учета долгосрочных потребностей практики в научных исследованиях можно обратиться к концепции тройной спирали Г. Ицковица [Ицковиц, 2010].

В основе концепции лежит модель взаимодействия государства, университета и бизнеса, названная автором тройной спиралью. Автор концепции на многочисленных примерах создания инновационных центров в США, Швеции и Бразилии демонстрирует, что в разных странах и географических регионах одного государства наблюдаются различные формы реализации модели тройной спирали. Но общим моментом во всех рассмотренных примерах является формирование условий непрерывного создания в регионе новых фирм, реализующих идеи, разработанные в стенах университетов. Ключевым условием появления новых фирм автор считает создание в регионе университета нового типа — предпринимательского. Прослеживая исторические причины успешности двух эталонных предпринимательских вузов, Массачусетского технологического института и Стэнфордского университета, можно сделать вывод, что видимые и активно воспроизводимые в других университетах структуры, такие как центры трансфера технологий, инкубаторы и технопарки, являются вторичными по отношению к концепции предпринимательских университетов. Университет нового типа не просто создает региональный пул экспертов — преподавателей университета, но инициирует передачу своих разработок в региональную промышленность, способ-

## Перечень универсальных навыков для исследователей

Источник: [Transferable Skills..., 2012, p. 20]

| Категория навыков             | Перечень универсальных навыков  |
|-------------------------------|---|
| Когнитивные способности       | <ul style="list-style-type: none"> <li>• Креативность (творческий подход) и абстрактное мышление.</li> <li>• Решение задач.</li> </ul>  |
| Навыки коммуникаций           | <ul style="list-style-type: none"> <li>• Письменные и устные навыки общения и подачи информации.</li> <li>• Навыки общения и ведения диалога с неспециалистами (связи с общественностью).</li> <li>• Навыки преподавания / группового обучения.</li> <li>• Научный подход к формированию политики организации и законотворчеству.</li> </ul>        |
| Предпринимательские навыки    | <ul style="list-style-type: none"> <li>• Бизнес-навыки (основы предпринимательской деятельности, инновации, коммерциализация новых товаров, интеллектуальная собственность и патентное дело).</li> </ul>  |
| Межличностные навыки          | <ul style="list-style-type: none"> <li>• Групповая деятельность / работа в команде.</li> <li>• Навыки репетиторства / индивидуального обучения.</li> <li>• Ведение переговоров.</li> <li>• Налаживание связей (установление контактов) и неформальное общение.</li> </ul>   |
| Организационные навыки        | <ul style="list-style-type: none"> <li>• Навыки ведения проектов и планирование рабочего времени.</li> <li>• Планирование карьеры.</li> </ul>   |
| Исследовательские компетенции | <ul style="list-style-type: none"> <li>• Навыки планирования и управления исследовательской работой.</li> <li>• Знакомство с методами и технологиями из областей научного знания, лежащих вне тематики квалификационной работы.</li> <li>• Научная этика и принципы научной честности.</li> <li>• Написание заявок на получение грантов.</li> </ul> |

ствуя созданию новых фирм из числа своих ученых и выпускников, если в регионе нет фирм, готовых реализовать идеи. Автором концепции университета как инициатора передачи технологий в промышленность был Уильям Бартон Роджер, основатель и первый президент Массачусетского технологического института, изложивший свои идеи в докладе 1846 г. Спустя сто лет данная концепция была реализована в Стэнфордском университете благодаря Фредерику Терману, который, работая во время войны в Массачусетском технологическом инсти-

туте, мог наблюдать эффективность концепции в военных разработках.

Государство в предложенной модели выступает не просто как гарант правил игры, но и как источник инвестиций с высоким риском и горизонтом получения результатов десять и более лет. По мнению Г. Ицковица, гарантией того, что инвестиции со столь долгим сроком появления первых результатов будут эффективны, является то, что каждый элемент в тройной спирали обеспечивает процесс уточнения целей исследований с учетом потребностей практики, ко-

торые могут возникнуть в будущем. Модель тройной спирали демонстрирует, что идеи исследований в университетах возникают в результате общения представителей региональной промышленности и вузовской науки в рамках ассоциаций, поддерживаемых региональными властями, но финансирование идеи получают только в том случае, если проходят экспертизу на предмет возможности практического применения на горизонте десятилетий.

## Заключение

Методы визуального представления наукометрической информации позволяют сформировать начальные гипотезы о закономерностях при анализе наукометрической информации и представлении конечных результатов анализа. Обнаруженные на диаграммах закономерности обязательно

требуют подтверждения в рамках дальнейших исследований с опорой на методы количественного анализа, представленные в главах 2 и 3 данной книги. Показана связь между приемами компактного представления результатов наукометрического анализа и логикой развития и совершенствования научного процесса. Проведенный в главе анализ свидетельствует, что при принятии решений о векторе развития науки на макроуровне нельзя опираться исключительно на методы наукометрии, привлечение экспертного мнения принципиально важно для получения адекватного результата. Развитие новых научных направлений на уровне организаций зависит в большей степени от наличия и уровня развития человеческого капитала и в меньшей степени от определения перспективных направлений развития науки и техники.

Желязны Дж. Говори на языке диаграмм. Пособие по визуальным коммуникациям. М.: Манн, Иванов и Фербер, 2009. 320 с.

Ицкович Г. Тройная спираль: университеты — предприятия — государство: инновации в действии. Томск: Изд-во Томского гос. ун-та систем управления, 2010. 237 с.

Латур Б. Наука в действии: Прагматический поворот. СПб.: Изд-во Европейского университета в Санкт-Петербурге, 2013. 416 с. [Оригинал: Latour B. Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge, Mass.: Harvard University Press, 1988. 288 p.]

Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми образами. М.: Манн, Иванов и Фербер, 2013. 352 с.

An approach to discovering new technology opportunities: Keyword-based patent map approach / S. Lee, B. Yoon, Y. Park // *Technovation*. 2009. Vol. 29. № 6–7. P. 481–497.

Börner, K., Polley, D.E. Visual Insights: A Practical Guide to Making Sense of Data. Cambridge, Massachusetts: The MIT Press, 2014. 312 p.

CitNetExplorer - Analyzing citation patterns in scientific literature. [2014]. URL: <http://www.citnetexplorer.nl/Home> (дата обращения: 22.07.2014).

Clickstream Data Yields High-Resolution Maps of Science / J. Bollen et al. // *PLoS ONE*. 2009. Vol. 4. № 3. P. e4803.

Collins R. The Sociology of Philosophies: A Global Theory of Intellectual Change. Belknap Press of Harvard University Press, 2000. 1098 p. [Русский перевод: Коллинз Р. Социология философий: глобальная теория интеллектуального изменения. Новосибирск: Сибирский хронограф, 2002. 1280 с.]

Design and Update of a Classification System: The UCSD Map of Science / K. Borner et al. // *PLoS ONE*. 2012. Vol. 7. № 7. P. e39464. [Материалы классификации опубликованы по адресу <http://sci.cns.iu.edu/ucsdmap/>]

Eck N. J. van, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping // *Scientometrics*. 2010. Vol. 84. № 2. P. 523–538.

Eck N. J. van, Waltman L. CitNetExplorer: A new software tool for analyzing and visualizing citation networks // *Journal of Informetrics*. 2014. Vol. 8. № 4. P. 802–823.

HistCite / ThomsonReuters. [2012]. URL: <http://interest.science.thomsonreuters.com/forms/HistCite/> (дата обращения: 22.07.2014).

IVMOOC: Information Visualization [Открытый онлайн курс]. [2014]. URL: <http://ivmooc.cns.iu.edu/> (дата обращения: 21.06.2014).

*Lotka A. J.* The Frequency Distribution of Scientific Productivity // *Journal of the Washington Academy of Sciences*. 1926. № 16. P. 317–323.

Mapping research specialties / S. A. Morris, B. Van der Veer Martens // *Annual Review of Information Science and Technology*. 2008. Vol. 42. № 1. P. 213–295.

Pajek / Vladimir Batagelj, Andrej Mrvar. 1996.

Participatory visualization with wordle / F. B. Viegas, M. Wattenberg, J. Feinberg // *Visualization and Computer Graphics, IEEE Transactions on*. 2009. Vol. 15. № 6. P. 1137–1144.

Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004 / R. Fontana, A. Nuvolari, H. Shimizu, A. Vezzulli // *Research Policy*. Ser. Economics, innovation and history: Perspectives in honour of Nick von Tunzelmann. 2013. Vol. 42. № 10. P. 1780–1792.

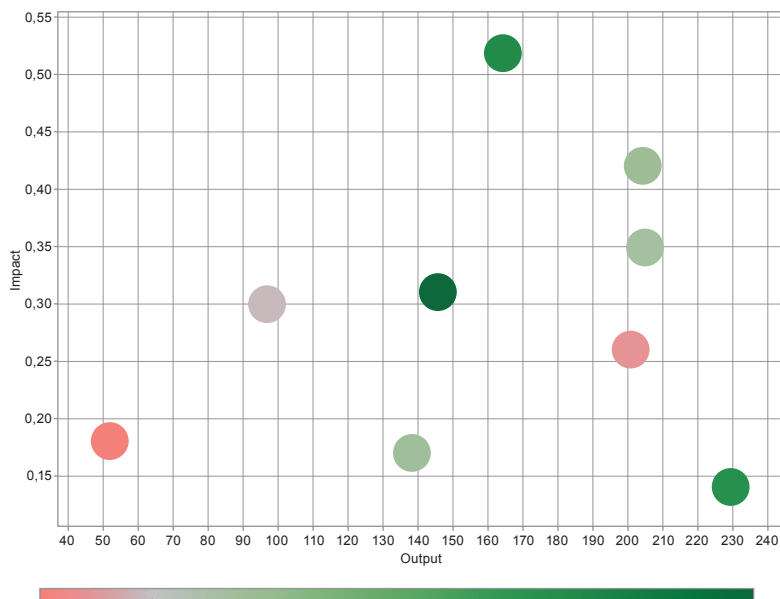
Sci<sup>2</sup> Tool : A Tool for Science of Science Research and Practice. [2010]. URL: <https://sci2.cns.iu.edu/user/index.php> (дата обращения: 23.07.2014).

Science mapping software tools: Review, analysis, and cooperative study among tools / M. J. Cobo, A. G. Lopez-Herrera, E. Herrera-Viedma, F. Herrera // *Journal of the American Society for Information Science and Technology*. 2011. Vol. 62. № 7. P. 1382–1402.

Tracking scientific development and collaborations – The case of 25 Asian countries / Henk Moed, Gali Halevi // *Research Trends*. 2014. № 38. P. 25–30.

Transferable Skills Training for Researchers / OECD. Paris: Organisation for Economic Co-operation and Development, 2012. 151 p.

VOSviewer // Centre for Science and Technology Studies, Leiden University, 2013. URL: <http://www.vosviewer.com>. (дата обращения: 25.06.2014).



а)

| Предметная область | Число публикаций | Нормализованная цитируемость публикаций | Рост числа публикаций в 2009-2013 гг. по сравнению с 2004-2008 гг. |
|--------------------|------------------|---|--|
| Subject_1          | 205              | 0,35                                    | 13,50%   |
| Subject_2          | 201              | 0,26                                    | -12,10%  |
| Subject_3          | 204              | 0,42                                    | 17,00%   |
| Subject_4          | 229              | 0,14                                    | 72,60%   |
| Subject_5          | 97               | 0,3                                     | -2,00%   |
| Subject_6          | 138              | 0,17                                    | 15,60%   |
| Subject_7          | 146              | 0,31                                    | 110,60%  |
| Subject_8          | 10               | 0,83                                    | 800,00%  |
| Subject_9          | 52               | 0,18                                    | -20,70%  |
| Subject_10         | 164              | 0,52                                    | 78,00%   |

б)

Рис. 36. Представление сравнительных количественных и качественных характеристик публикаций по предметным областям (см. пояснение на с. 165–167):

а) графическое представление предметных областей на пузырьковой диаграмме, где по горизонтали представлено количество публикаций (Output), а по вертикали – нормализованная цитируемость публикаций (Impact, см. раздел 2.3); цветом закодирована динамика роста за десять лет; предметная область Subject\_8 исключена из диаграммы по причине малого числа публикаций по сравнению с остальными областями;

б) исходные данные, показанные на предыдущей диаграмме; цветом закодирован диапазон значений по столбцам

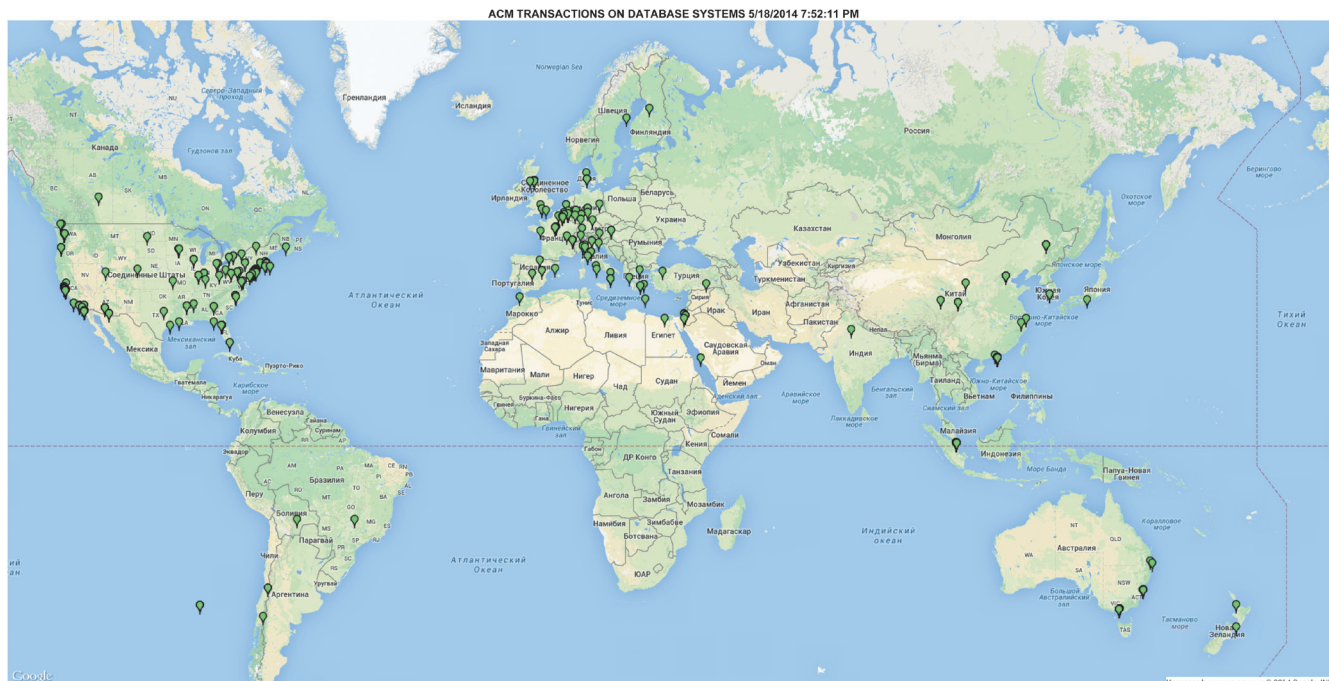


Рис. 37. Географическое распределение авторов ACM Transactions on Database Systems (TODS) 2008–2014 гг.  
Пояснение см. на с. 168–169

1926  
1928  
1930  
1932  
1934  
1936  
1938  
1940  
1942  
1944  
1946  
1948  
1950  
1952  
1954  
1956  
1958  
1960  
1962  
1964  
1966  
1968  
1970  
1972  
1974  
1976  
1978  
1980  
1982  
1984  
1986  
1988  
1990  
1992  
1994  
1996  
1998  
2000  
2002  
2004  
2006  
2008  
2010  
2012  
2014

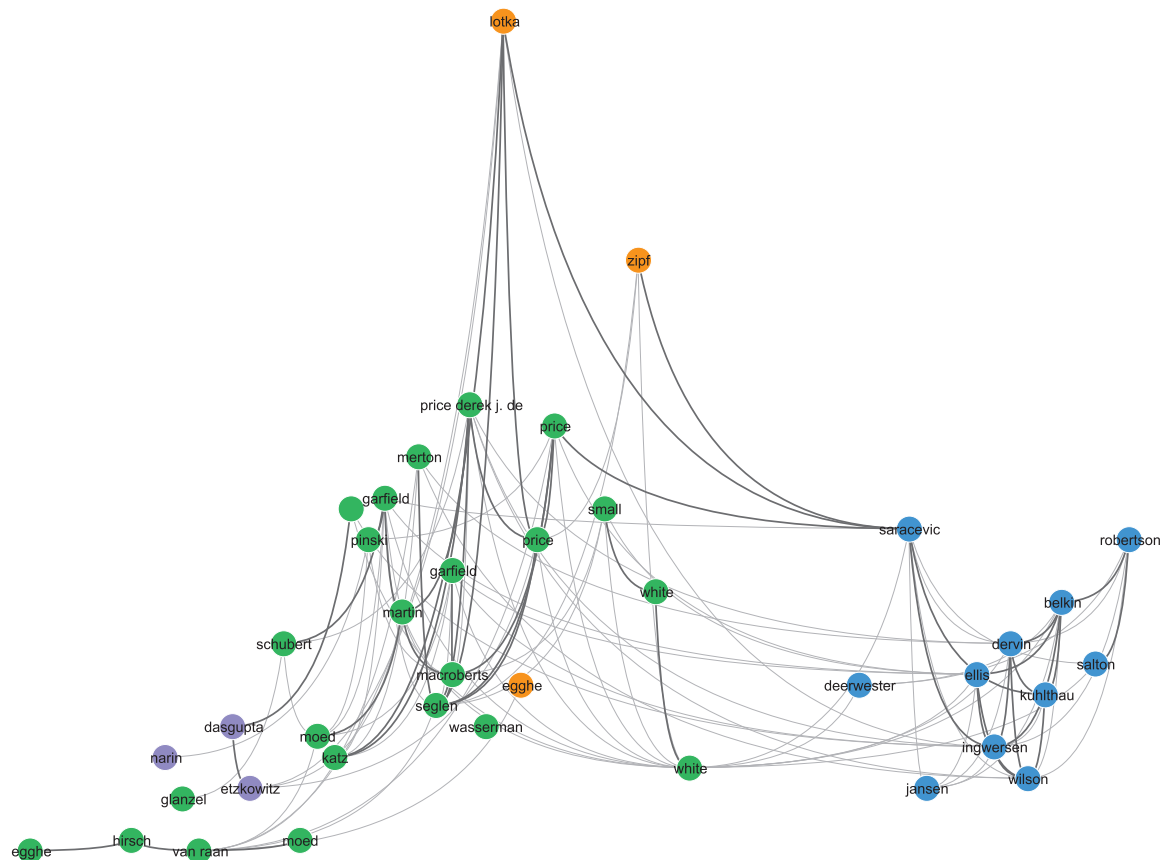


Рис. 38. Связи между кластерами публикаций, прямо или косвенно цитирующих работу [Lotka, 1926] в массиве публикаций за период 1945–2014 гг. Представлено 4284 публикации и 28482 ссылки. В качестве меток кластеров используется фамилия первого автора. Исходный массив публикаций описан в тексте.

Пояснение см. на с. 170





Рис. 39. Сеть связей соавторства публикаций четырех авторов: A.-L. Barabasi, A. Vespignani, E. Garfield, S. Wasserman (361 запись, период 1955–2007 гг., источник записей – Web of Science CC).

Построено по методике из [Börner, Polley, 2014, p. 204–206].

Пояснение см. на с. 171



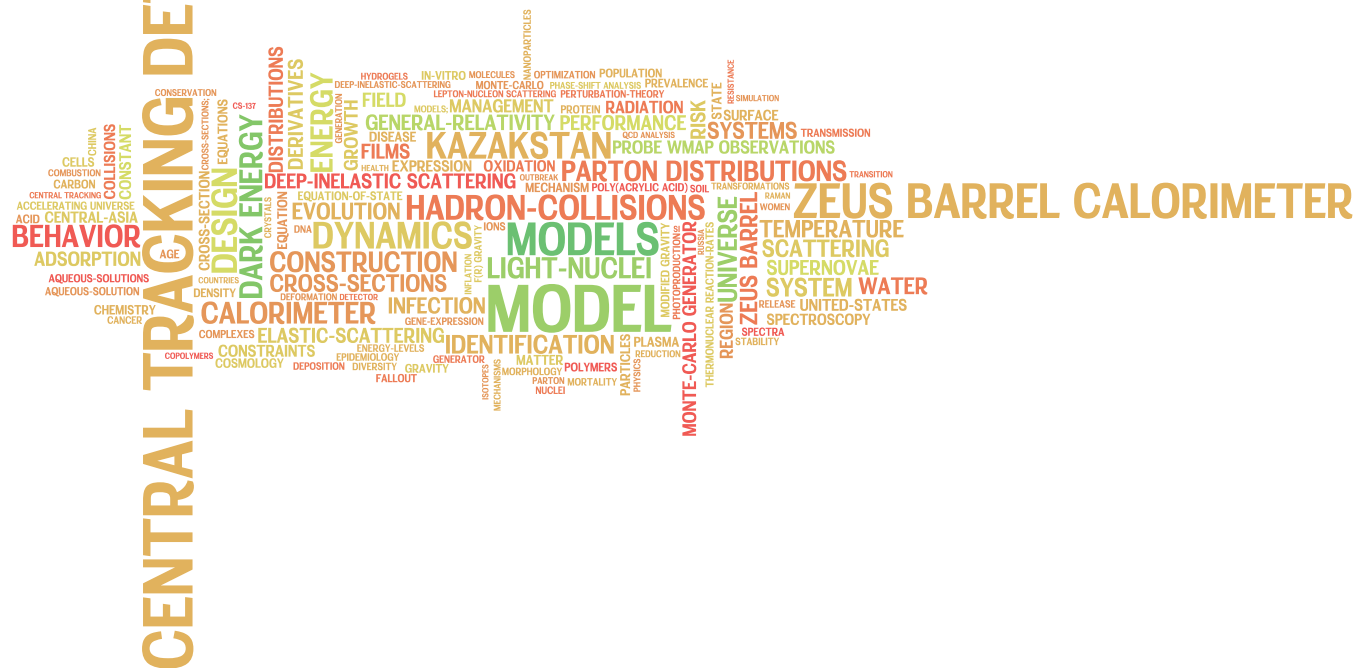


Рис. 41. Облако ключевых слов, указанных в публикациях авторов из Казахстана. Цветом показано изменение числа ключевых слов за 2009–2014 гг. по сравнению с 2004–2008 гг. Показано 140 ключевых слов из поля KeyWords Plus, встречавшихся более 10 раз (20 % от всех указанных в публикациях). Проанализировано 4737 публикаций за 2004–2014 гг., где хотя бы один автор указал в аффилиации Казахстан. Источник записей: Web of Science CC. Пояснение см. на с. 173

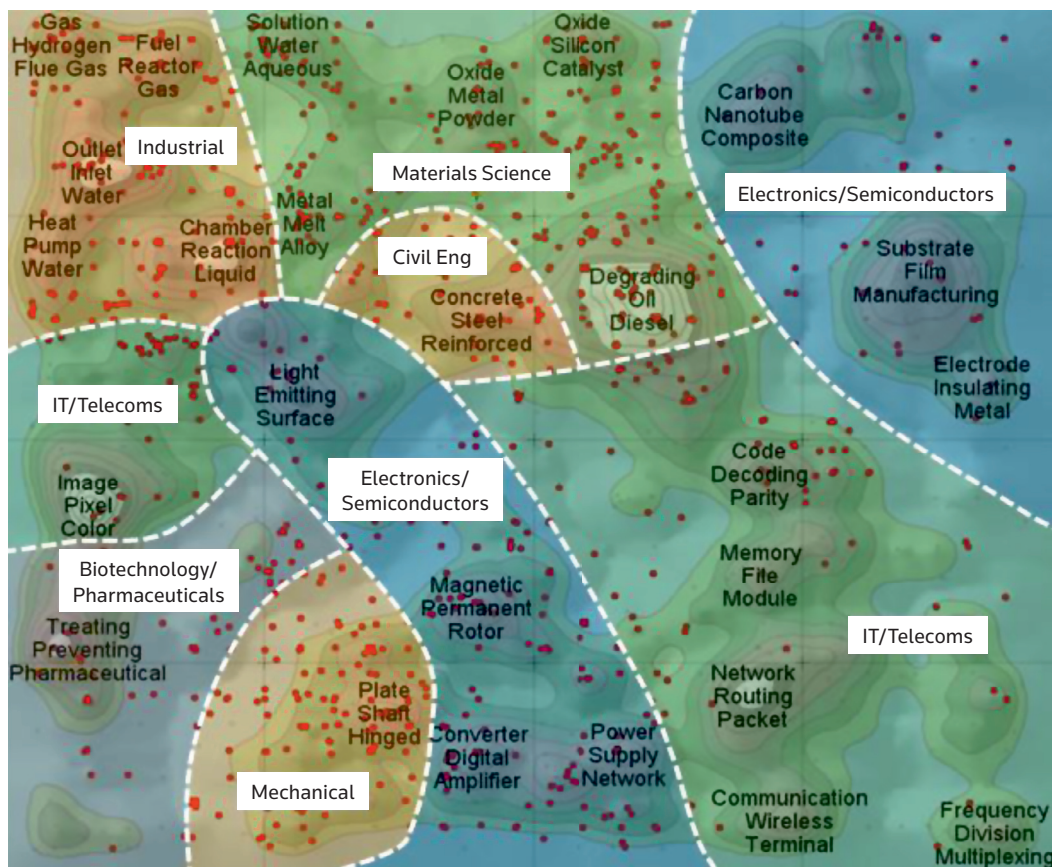


Рис. 42. Патентный ландшафт УрФУ (Theme Scape Map). Точками отмечены патенты, полученные УрФУ за период 1993–2001 гг. Ландшафт построен на основании патентных семейств, полученных за указанный период времени 8 университетами, выбранными в качестве бенчмарков, и 20 институтами УрО РАН. Инструмент – ThomsonInnovation. Источник: Отчет НИР для разработки плана развития науки в УрФУ, первый этап. Thomson Reuters, 2013.

Пояснение см. на с. 174

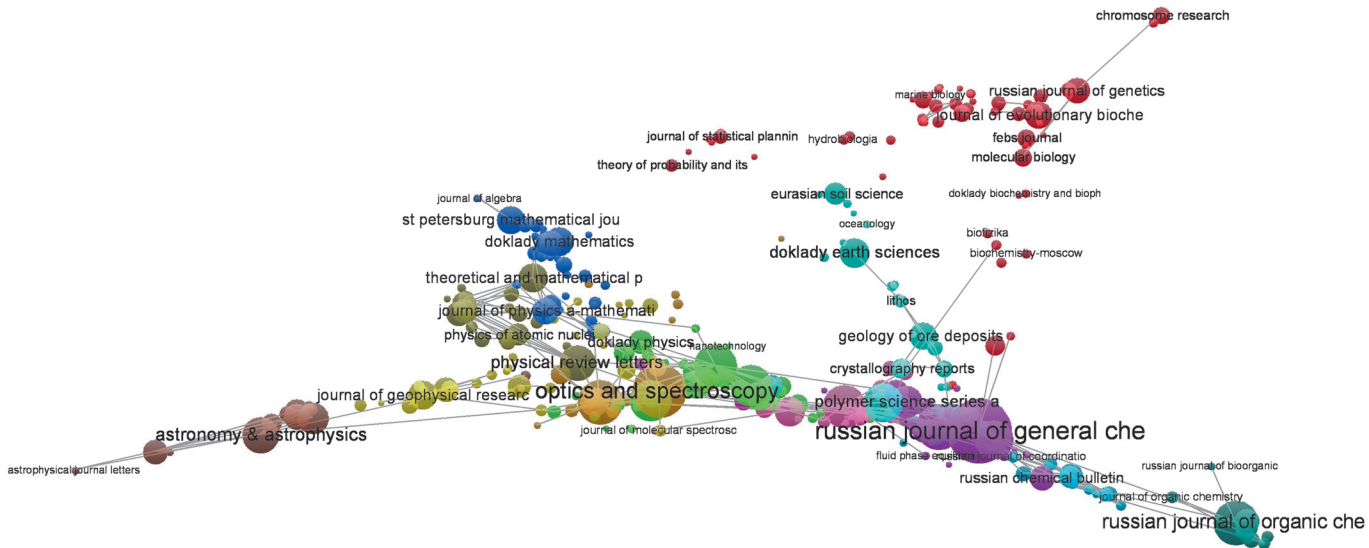


Рис. 43. Визуализация методом «библиографического сочетания» по журналам массива публикаций авторов СПбГУ (10785 записей, период 2004–2013 гг., источник записей – Web of Science CC, построено с использованием VOSviewer). Цветом закодированы тематические категории; площадь круга отражает число публикаций в источнике; расположение и линии связи указывают на силу связи публикаций, цитирующих работы, опубликованные в одном и том же журнале.

Пояснение см. на с. 175–176



Рис. 44. Карта науки, построенная методом ко-цитирования для 5000 базовых научных журналов (публикации за период 2007 г., анализ цитирований за 1997–2006 гг., источник записей – Web of Science CC, построено с использованием VOSviewer). Цветом закодированы тематические категории; площадь круга отражает число публикаций в источнике; расположение указывает на силу связи цитирования одинакового набора публикаций; описание метода построения представлено в [Eck, Waltman, 2010]. Обработанные данные для построения карты включены в состав поставки VOSviewer.

Пояснение см. на с. 176

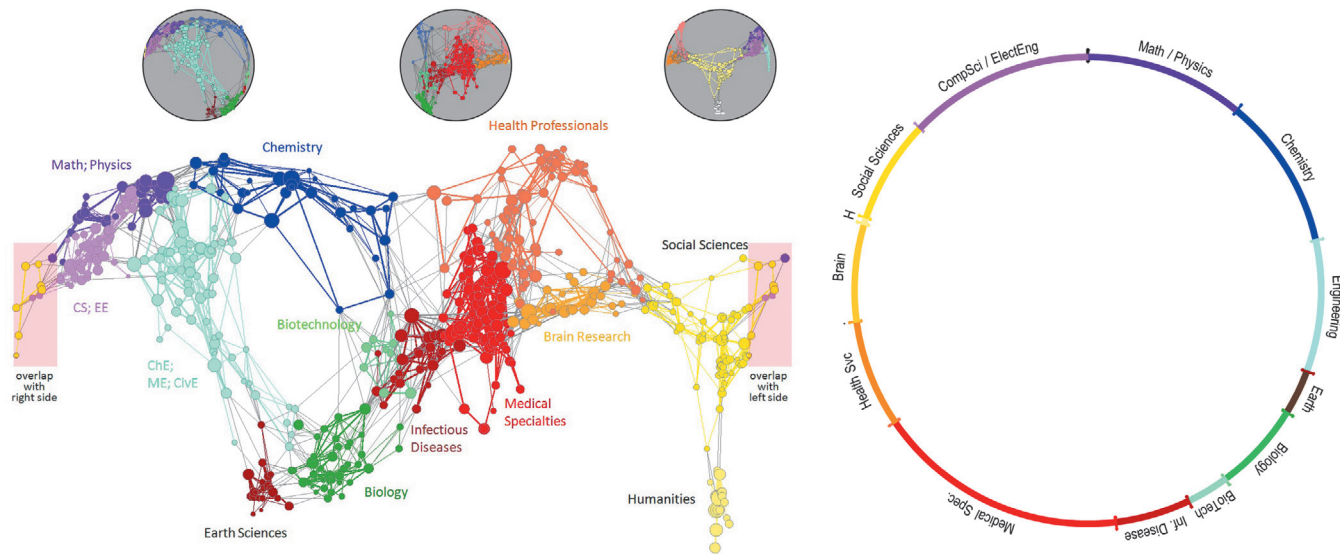


Рис. 45. Визуализация карты науки UCSD (2001–2010 гг.). Слева – отображение карты науки в двумерной проекции Меркатора (левая сторона карты имеет продолжение справа), сверху – три врезки с отображением на 3D-сферу, справа – отображение на одномерное колесо. Источник: [Design and Update..., 2012]. Карта опубликована под лицензией Creative Commons, права на классификацию и карту принадлежат University of California, San Diego.

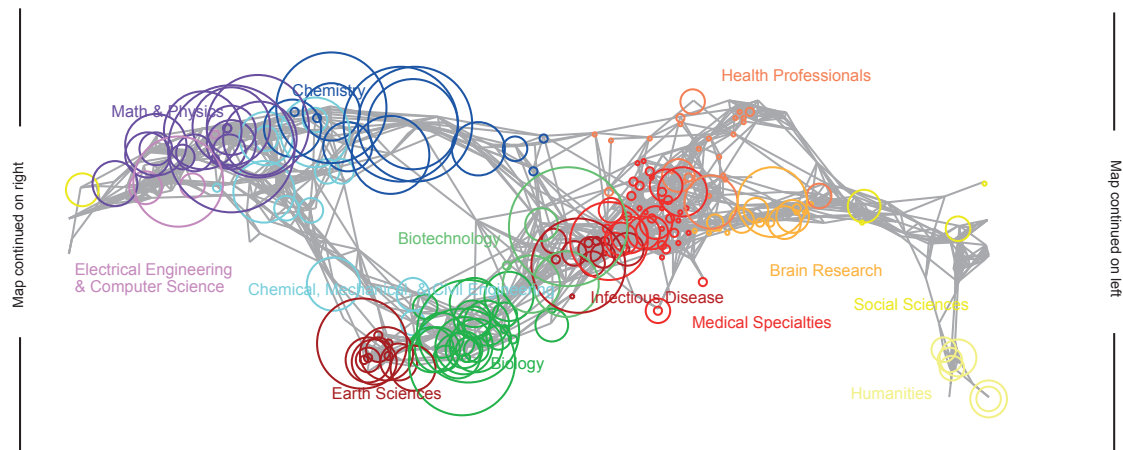
Пояснение см. на с. 177

## Topical Visualization

Generated from 10785 Unique ISI Records

154 out of 1 595 publications were mapped to 202 subdisciplines and 13 disciplines.

19, 2014 | 10:46 PM YEKT

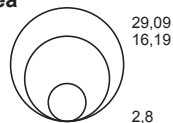


2008 The Regents of the University of California and SciTech Strategies.  
Map updated by SciTech Strategies, OST, and CNS in 2011.

### Legend

Circle area: Fractional Journal Count  
Unclassified = 1441  
Minimum = 0  
Maximum = 36  
Color: Discipline  
See end of PDF for color legend.

### Area



CNS ([cns.iu.edu](http://cns.iu.edu))

### How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

Рис. 46. Визуализация массива публикаций авторов СПбГУ на карте классификации UCSD (10785 записей, период – 2004–2013 гг., источник записей – Web of Science CC, построено с использованием Sci<sup>2</sup> Tool). Цветом закодированы категории; площадь круга отражает число публикаций в журналах подкатегории; расположение соответствует карте подкатегорий журналов, представленной на рис. 45. Описание метода построения представлено в [Börner, Polley, 2014, p. 140–141].

Пояснение см. на с. 177